Some Properties of $[\mathrm{tr}(Q^{2p})]^{1/2p}$ with Application to Linear Minimax Estimation

J. Lauterbach
Department of Economics
University of Hannover
3000 Hannover, West Germany

and

P. Stahlecker
Department of Law and Economics
University of Mainz
6500 Mainz, West Germany

Submitted by F. Pukelsheim

ABSTRACT

A nondifferentiable minimization problem is considered which occurs in linear minimax estimation. This problem is solved by replacing the nondifferentiable maximal eigenvalue of a real nonnegative definite matrix Q with $[\mathrm{tr}(Q^{2p})]^{1/2p}$ $(p \in \mathbb{N})$. It is shown that any descent algorithm with inexact step-length rule can be used to obtain linear minimax estimators for the parameter vector of a parameter-restricted linear model.

1. INTRODUCTION

Since the pioneer study of Kuks and Olman [12], the problem of obtaining linear minimax estimators in a parameter-restricted linear model has received considerable interest (cf. Läuter [13], Hoffmann [10], Toutenburg [23], Pilz [18, 19], Stahlecker and Trenkler [22], Drygas [4], Gaffke and Heiligers [6], and others). For a particular class of prior restrictions to be specified subsequently (see Section 6) it can be shown that minimax linear estimation leads to the problem of solving the nondifferentiable minimization problem

$$f(C) \to \min, \tag{1}$$

655 Avenue of the Americas, New York, NY 10010 0024-3795/90/$3.50

where

$$f:\begin{cases}\mathbb{R}^{r\times n}\to\mathbb{R},\\ C\to\rho\,\mathrm{tr}(VC'AC)+\lambda_{\max}(Q(C)),\end{cases}\tag{2}$$

and

$$Q:\begin{cases}\mathbb{R}^{r\times n}\to\mathbb{R}^{k\times k}_{\geqslant},\\ C\to(CW-S)'A(CW-S).\end{cases}\tag{3}$$

Here A, V, W, and S are known $r\times r$, $n\times n$, $n\times k$, and $r\times k$ real matrices, respectively, where A is nonnegative definite (n.n.d.), V is positive definite (p.d.), the scalar ρ is assumed to be positive, $\mathrm{tr}(B)$ denotes the trace of the matrix B, and $\lambda_{\max}(Q(C))$ denotes the maximal eigenvalue of the n.n.d. matrix $Q(C)$.

Without loss of generality we may assume $V=I$. Now let $A=K'K$ be a rank factorization of A. Then for each $C\in\mathbb{R}^{r\times n}$

$$\begin{aligned}f(C)&=\rho\,\mathrm{tr}(C'AC)+\lambda_{\max}(Q(C))\\&=\rho\,\mathrm{tr}((KC)'KC)+\lambda_{\max}((KCW-KS)'(KCW-KS)),\end{aligned}$$

and for each $D\in\mathbb{R}^{m\times n}$ $[m:=\mathrm{rank}(A)]$

$$\begin{aligned}&\rho\,\mathrm{tr}(D'D)+\lambda_{\max}((DW-KS)'(DW-KS))\\&\quad=\rho\,\mathrm{tr}\Big(\big[K'(KK')^{-1}D\big]'A\big[K'(KK')^{-1}D\big]\Big)\\&\qquad+\lambda_{\max}\Big(\big[K'(KK')^{-1}DW-S\big]'A\big[K'(KK')^{-1}DW-S\big]\Big)\\&\quad=f\big(K'(KK')^{-1}D\big).\end{aligned}$$

Hence we may consider only the case $A=I$. Furthermore, we assume subsequently

$$W\neq 0\tag{4}$$

and

$$S\neq 0,\tag{5}$$

since otherwise we would obtain the trivial solution $C_*=0$.

Since the first term in (2) is uniformly convex and $\lambda_{\max}(Q(\cdot))$ is convex, the function f is uniformly convex. This implies that the level set

$$\mathscr{W}(C_0) = \left\{C \in \mathbb{R}^{r\times n} \middle| f(C) \leqslant f(C_0)\right\}$$

is compact for each $C_0 \in \mathbb{R}^{r\times n}$. Hence there exists a solution C_* of the problem (1), which is unique by the uniform convexity of f. This solution C_*, however, can only be determined explicitly in special cases (see for instance Läuter [13], Hoffmann [10], Pilz [18], and Gaffke and Heiligers [6]).

Instead of solving (1) directly by methods for nondifferentiable optimization, we could replace $\lambda_{\max}(Q(C))$ by $\|Q(C)^p\|^{1/p}$, where $\|\cdot\|$ denotes any matrix norm and $p \in \mathbb{N}$. Since for any $M \in \mathbb{R}^{k\times k}$

$$\lim_{p\to\infty} \|M^p\|^{1/p} = \sigma(M), \tag{6}$$

where

$$\sigma(M) = \max\{|\lambda| \,|\, \lambda \text{ is an eigenvalue of } M\} \tag{7}$$

is the spectral radius of M (cf. for instance Horn and Johnson [11, p. 299]), and since $Q(C)$ is n.n.d., it follows that for any $C \in \mathbb{R}^{r\times n}$

$$\lim_{p\to\infty} \left\|Q(C)^p\right\|^{1/p} = \lambda_{\max}(Q(C)). \tag{8}$$

In particular, choosing the differentiable Frobenius norm $\|M\|_F = [\operatorname{tr}(M'M)]^{1/2}$, we obtain

$$\lim_{p\to\infty} \left[\operatorname{tr}\left(Q(C)^{2p}\right)\right]^{1/2p} = \lambda_{\max}(Q(C)). \tag{9}$$

Thus, defining $f_p : \mathbb{R}^{r\times n} \to \mathbb{R}$ by

$$f_p(C) := \rho \operatorname{tr}(VC'AC) + \left[\operatorname{tr}\left(Q(C)^{2p}\right)\right]^{1/2p}, \tag{10}$$

we get for any fixed $C \in \mathbb{R}^{r\times n}$

$$\lim_{p\to\infty} f_p(C) = f(C). \tag{11}$$

There exists a (unique) solution C_p of the problem

$$f_p(C) \to \min \tag{12}$$

(see Section 4). Moreover, for any $C \in \mathbb{R}^{r \times n}$ the inequalities

$$f_p(C_p) - f(C_*) \leqslant (k^{1/2p} - 1) f(C) \tag{13}$$

and

$$\|C_p - C_*\|^2 \leqslant \frac{1}{\rho}(k^{1/2p} - 1) f(C) \tag{14}$$

are satisfied (cf. Stahlecker and Lauterbach [21]). Thus, we obtain

$$\lim_{p \to \infty} C_p = C_*$$

and

$$\lim_{p \to \infty} f_p(C_p) = f(C_*).$$

From these observations the following idea can be motivated. For an increasing sequence $(p_j)_{j \in \mathbb{N}}$, where $p_j \in \mathbb{N}$, solve the auxiliary problem

$$f_{p_j}(C) \to \min \tag{15}$$

by some numerical method, where a numerical solution $\overline{C}_{p_{j-1}}$ may be used as a starting point. Does this procedure lead to a reasonable approximation for C_*?

Actually (see Theorem 2) we may determine C_* and $f(C_*)$ up to any degree of precision, for which a wide class of methods for unconstrained optimization including second-order methods (i.e. quasi-Newton methods) can be used to solve the subproblems (15). This generalizes our previous results where we considered the steepest-descent algorithm (cf. Stahlecker and Lauterbach [21]). As an application, a minimax problem with affine linear restrictions, recently investigated by Stahlecker and Trenkler [22] and Drygas [4], can be resolved completely.

Subsequently we use the following notation:

$$\begin{aligned}
\mathbb{R}^{k\times k}_{\mathscr{S}} &= \text{the linear space of real symmetric } k\times k \text{ matrices},\\
\mathbb{R}^{k\times k}_{\geqslant} &= \text{the set of all n.n.d. matrices } Z\in\mathbb{R}^{k\times k}_{\mathscr{S}},\\
Z^{+} &= \text{the Moore-Penrose inverse of the matrix } Z,\\
\mathscr{L}(\mathscr{U},\mathscr{V}) &= \text{the space of linear operators } G:\mathscr{U}\to\mathscr{V},\\
\nabla h[x]\ (\nabla^2 h[x]) &= \text{the first (second) Fréchet derivative of the function } h,\\
\langle R,S\rangle := \operatorname{tr}(R'S) &= \text{the inner product on } \mathbb{R}^{r\times n},\\
\|R\| := [\operatorname{tr}(R'R)]^{1/2} &= \text{the Frobenius norm of the matrix } R.
\end{aligned}$$

2. SOME PROPERTIES OF $[\operatorname{tr}(Q^{2p})]^{1/2p}$

We start from the functional

$$h_p:\begin{cases}\mathbb{R}^{k\times k}_{\mathscr{S}}\to\mathbb{R},\\ Q\to\left[\operatorname{tr}\left(Q^{2p}\right)\right]^{1/2p}\end{cases} \tag{16}$$

for fixed $p\in\mathbb{N}$. First observe that h_p has the following properties:

(i) h_p is convex,
(ii) h_p is linear homogeneous, i.e., $h_p(\lambda Q)=|\lambda|h_p(Q)$ for all $\lambda\in\mathbb{R}$, and
(iii) h_p is subadditive, i.e., $h_p(Q+Z)\leqslant h_p(Q)+h_p(Z)$ for all $Q,Z\in\mathbb{R}^{k\times k}_{\mathscr{S}}$.

Statement (i) follows by applying results of Marshall and Olkin [15, pp. 96, 245]), (ii) is obvious, and (iii) follows from (i) and (ii). Note that (ii) and (iii) mean that h_p is sublinear.

We are now interested in formulas for the first and second derivative.

PROPOSITION 1. *The functional h_p is twice Fréchet-differentiable on $\mathbb{R}^{k\times k}_{\mathscr{S}}\setminus\{0\}$, and*

(i) $\nabla h_p[Q]\in\mathscr{L}(\mathbb{R}^{k\times k},\mathbb{R})$ *is given by*

$$\nabla h_p[Q](Z)=s_p^{(1)}(Q)\operatorname{tr}(Q^{2p-1}Z), \tag{17}$$

(ii) $\nabla^2 h_p[Q] \in \mathscr{L}(\mathbb{R}^{k\times k}\times\mathbb{R}^{k\times k},\mathbb{R})$ *is given by*

$$\nabla^2 h_p[Q](Z_1,Z_2) = (1-2p)s_p^{(2)}(Q)\mathrm{tr}\left(Q^{2p-1}Z_2\right)\mathrm{tr}\left(Q^{2p-1}Z_1\right) + s_p^{(1)}(Q)\mathrm{tr}\left(\sum_{i=1}^{2p-1} Q^{i-1}Z_2Q^{2p-i-1}Z_1\right), \tag{18}$$

where $s_p^{(i)}(Q) := [\mathrm{tr}(Q^{2p})]^{(1/2p)-i}$, $i=1,2$.

Proof. (i): Omitted (see for instance Magnus and Neudecker [14, p. 177 and p. 183]).

(ii): Consider the gradient H_p of h_p defined by $\nabla h_p[Q](Z) = \langle H_p(Q), Z\rangle$, i.e.

$$H_p: \begin{cases} \mathbb{R}^{k\times k}_{\mathscr{S}} \to \mathbb{R}^{k\times k}, \\ Q \to s_p^{(1)}(Q)Q^{2p-1}. \end{cases} \tag{19}$$

Then $\nabla^2 h_p[Q](Z_1, Z_2) = \mathrm{tr}(\nabla H_p[Q](Z_2)Z_1)$. Since $H_p(Q) = s_p^{(1)}(Q)V_{2p-1}(Q)$, where

$$V_{2p}: \begin{cases} \mathbb{R}^{k\times k}_{\mathscr{S}} \to \mathbb{R}^{k\times k}_{\geqslant}, \\ Q \to Q^{2p}, \end{cases}$$

it follows by the product rule (Nashed [16, p. 165]) that

$$\begin{aligned}\nabla H_p[Q](Z_2) &= \nabla s_p^{(1)}[Q](Z_2)V_{2p-1}(Q) + s_p^{(1)}(Q)\nabla V_{2p-1}[Q](Z_2) \\ &= (1-2p)s_p^{(2)}(Q)\mathrm{tr}\left(Q^{2p-1}Z_2\right)Q^{2p-1} \\ &\quad + s_p^{(1)}(Q)\left(\sum_{i=1}^{2p-1} Q^{i-1}Z_2Q^{2p-i-1}\right).\end{aligned}$$

Thus, we get

$$\nabla^2 h_p[Q](Z_1,Z_2) = (1-2p)s_p^{(2)}(Q)\mathrm{tr}\left(Q^{2p-1}Z_2\right)\mathrm{tr}\left(Q^{2p-1}Z_1\right) + s_p^{(1)}(Q)\mathrm{tr}\left(\sum_{i=1}^{2p-1} Q^{i-1}Z_2Q^{2p-i-1}Z_1\right). \quad \blacksquare$$

Since h_p is convex, the well-known inequalities (see for instance Daniel [2, p. 13])

$$h_p(Z) - h_p(Q) \geqslant \nabla h_p[Q](Z-Q) \tag{20}$$

and

$$\nabla^2 h_p[Q](Z,Z) \geqslant 0 \tag{21}$$

are satisfied for any $Q, Z \in \mathbb{R}^{k\times k}_{\mathscr{S}}$, $Q \neq 0$. Moreover, from (17) we obtain for $Q, Z \in \mathbb{R}^{k\times k}_{\geqslant}$, $Q \neq 0$

$$\nabla h_p[Q](Z) \geqslant 0. \tag{22}$$

PROPOSITION 2.

(i) *Let* $Q, Z \in \mathbb{R}^{k\times k}_{\geqslant}$ *such that* $Z - Q \in \mathbb{R}^{k\times k}_{\geqslant}$. *Then* $h_p(Z) \geqslant h_p(Q)$, *i.e.,* h_p *is isotone on* $\mathbb{R}^{k\times k}_{\geqslant}$.
(ii) *For any* $Q, Z \in \mathbb{R}^{k\times k}_{\geqslant}$, $Q \neq 0$, *we have* $\nabla h_p[Q](Z) \leqslant \operatorname{tr}(Z)$.
(iii) *For each* $Q \in \mathbb{R}^{k\times k}_{\mathscr{S}} \setminus \{0\}$ *and for each* $Z \in \mathbb{R}^{k\times k}$

$$\nabla^2 h_p[Q](Z,Z) \leqslant (2p-1)\left[\lambda_{\max}(Q^2)\right]^{-1/2}\|Z\|^2. \tag{23}$$

Proof. (i): This is obvious from (20) and (22).
(ii): Since h_p is subadditive, we obtain from (20)

$$\begin{aligned}\nabla h_p[Q](Z) + h_p(Q) &\leqslant h_p(Q+Z)\\ &\leqslant h_p(Q) + h_p(Z)\\ &\leqslant h_p(Q) + \operatorname{tr}(Z).\end{aligned}$$

(iii): By Proposition 1(ii)

$$\begin{aligned}\nabla^2 h_p[Q](Z,Z) &= (1-2p)s_p^{(2)}(Q)\left[\operatorname{tr}(Q^{2p-1}Z)\right]^2\\ &\quad + s_p^{(1)}(Q)\sum_{i=1}^{2p-1}\operatorname{tr}(Q^{i-1}ZQ^{2p-i-1}Z)\\ &\leqslant s_p^{(1)}(Q)\sum_{i=1}^{2p-1}\operatorname{tr}(Q^{i-1}ZQ^{2p-i-1}Z).\end{aligned}$$

Using the inequality $\operatorname{tr}(A'B) \leqslant \|A\| \|B\|$, we get

$$\begin{aligned} \operatorname{tr}\left(Q^{i-1} Z Q^{2p-i-1} Z\right) &\leqslant \left[\operatorname{tr}\left(Z' Q^{2(i-1)} Z\right)\right]^{1/2} \left[\operatorname{tr}\left(Z' Q^{2(2p-i-1)} Z\right)\right]^{1/2} \\ &\leqslant \left[\lambda_{\max}\left(Q^2\right)\right]^{(i-1)/2} \left[\lambda_{\max}\left(Q^2\right)\right]^{(2p-i-1)/2} \|Z\|^2 \\ &= \left[\lambda_{\max}\left(Q^2\right)\right]^{p-1} \|Z\|^2. \end{aligned}$$

Since $s_p^{(1)}(Q) \leqslant [\lambda_{\max}(Q^2)]^{(1/2)-p}$, it follows that

$$\nabla^2 h_p[Q](Z, Z) \leqslant (2p-1)\left[\lambda_{\max}\left(Q^2\right)\right]^{-1/2} \|Z\|^2. \qquad \blacksquare$$

Using the representation theorem for h_p (cf. Gaffke and Krafft [7, p. 619]),

$$h_p(Q) = \max\left\{\operatorname{tr}(QX) \,\middle|\, X \in \mathbb{R}_{\geqslant}^{k \times k},\ \left[\operatorname{tr}(Q^q)\right]^{1/q} = 1\right\},$$

where $Q \in \mathbb{R}_{\geqslant}^{k \times k}$, $1/2p + 1/q = 1$, we may also obtain Proposition 2(i) without differentiation. Obviously the maximal eigenvalue of $\nabla^2 h_p[Q]$ would yield the least upper bound in Proposition 2(iii). However, we will see subsequently that the inequality (23) can be used to derive a uniform upper bound for the hessian of our original function f_p. This will be of interest in our study of algorithmic solutions of (12) (see Section 5).

3. PROPERTIES OF $[\operatorname{tr}(Q(C)^{2p})]^{1/2p}$

Let now $Q: \mathbb{R}^{r \times n} \to \mathbb{R}_{\geqslant}^{k \times k}$ be an arbitrary matrix function satisfying the following assumptions:

(i) Q is twice Fréchet-differentiable.
(ii) For each $C, Z \in \mathbb{R}^{r \times n}$ the matrix $\nabla^2 Q[C](Z, Z)$ is n.n.d.
(iii) There exists $c > 0$ such that for each $C, Z \in \mathbb{R}^{r \times n}$

$$\left\|\nabla Q[C](Z)\right\|^2 \leqslant c \lambda_{\max}(Q(C)) \|Z\|^2. \tag{24}$$

(iv) There exists $d > 0$ such that for each $C, Z \in \mathbb{R}^{r \times n}$

$$\operatorname{tr}\left(\nabla^2 Q[C](Z, Z)\right) \leqslant d \|Z\|^2. \tag{25}$$

We will apply the results of Section 2 to the functional $g_p = h_p \circ Q$. First observe that by assumption (ii) the matrix function Q is convex with respect to the Löwner ordering on $\mathbb{R}^{k\times k}_{\geqslant}$ (for a proof see Lemma 1 of the appendix). Since h_p is convex and by Proposition 2(i) isotone on $\mathbb{R}^{k\times k}_{\geqslant}$, the following proposition is obvious.

PROPOSITION 3. *The functional g_p is convex on $\mathbb{R}^{r\times n}$.*

Let now

$$\mathcal{Q} := \left\{ C \in \mathbb{R}^{r\times n} \middle| Q(C) = 0 \right\}. \tag{26}$$

By applying the chain rule (cf. Nashed [16, p. 164]), i.e.

$$\nabla g_p[C](Z) = \nabla h_p[Q(C)](\nabla Q[C](Z)), \tag{27}$$

$$\nabla^2 g_p[C](Z_1, Z_2) = \nabla^2 h_p[Q(C)](\nabla Q[C](Z_1), \nabla Q[C](Z_2)) + \nabla h_p[Q(C)](\nabla^2 Q[C](Z_1, Z_2)), \tag{28}$$

we obtain the first and second Fréchet derivatives of $g_p = h_p \circ Q$.

PROPOSITION 4. *The function g_p has a first and a second Fréchet derivative on $\mathbb{R}^{r\times n} \setminus \mathcal{Q}$, and*

(i) *$\nabla g_p[C] \in \mathscr{L}(\mathbb{R}^{r\times n}, \mathbb{R})$ is given by*

$$\nabla g_p[C](Z) = s_p^{(1)}(C)\operatorname{tr}\left(Q(C)^{2p-1}\nabla Q[C](Z)\right), \tag{29}$$

(ii) *$\nabla^2 g_p[C] \in \mathscr{L}(\mathbb{R}^{r\times n} \times \mathbb{R}^{r\times n}, \mathbb{R})$ is given by*

$$\begin{aligned}\nabla^2 g_p[C](Z_1, Z_2) = s_p^{(1)}(C)\operatorname{tr}\Bigg(&\sum_{i=1}^{2p-1} Q(C)^{i-1}\nabla Q[C](Z_2) \\ &\times Q(C)^{2p-i-1}\nabla Q[C](Z_1)\Bigg) \\ &+ (1-2p)s_p^{(2)}(C)\operatorname{tr}\left(Q(C)^{2p-1}\nabla Q[C](Z_2)\right) \\ &\times \operatorname{tr}\left(Q(C)^{2p-1}\nabla Q[C](Z_1)\right) \\ &+ s_p^{(1)}(C)\operatorname{tr}\left(Q(C)^{2p-1}\nabla^2 Q[C](Z_1, Z_2)\right),\end{aligned} \tag{30}$$

where $s_p^{(i)}(C) = [\operatorname{tr}(Q(C)^{2p})]^{(1/2p)-i}$, $i = 1,2$.

Using the results of Section 2, we get a uniform lower and upper bound for the hessian of g_p.

PROPOSITION 5. *For each* $C \in \mathbb{R}^{r\times n}\setminus\mathcal{Q}$ *and* $Z \in \mathbb{R}^{r\times n}$ *we have*

$$0 \leqslant \nabla^2 g_p[C](Z,Z) \leqslant [(2p-1)c+d]\|Z\|^2. \tag{31}$$

Proof. The left-hand side of (31) is obvious from Proposition 3. By (28), Proposition 2(iii), and assumption (iii) we get

$$\begin{aligned}\nabla^2 g_p[C](Z,Z) &= \nabla^2 h_p[Q(C)](\nabla Q[C](Z), \nabla Q[C](Z))\\ &\quad + \nabla h_p[Q(C)](\nabla^2 Q[C](Z,Z))\\ &\leqslant (2p-1)[\lambda_{\max}(Q(C))]^{-1}\|\nabla Q[C](Z)\|^2\\ &\quad + \nabla h_p[Q(C)](\nabla^2 Q[C](Z,Z))\\ &\leqslant (2p-1)c\|Z\|^2 + \nabla h_p[Q(C)](\nabla^2 Q[C](Z,Z)).\end{aligned}$$

Furthermore, by Proposition 2(ii) we have

$$\nabla h_p[Q(C)](\nabla^2 Q[C](Z,Z)) \leqslant \operatorname{tr}(\nabla^2 Q[C](Z,Z)),$$

and hence by assumption (iv) it follows that

$$\nabla^2 g_p[C](Z,Z) \leqslant [(2p-1)c+d]\|Z\|^2. \qquad \blacksquare$$

Let us now consider the particular function

$$Q: \begin{cases} \mathbb{R}^{r\times n} \to \mathbb{R}^{k\times k}_{\geqslant} \\ C \to (CW-S)'(CW-S). \end{cases} \tag{32}$$

We will see that the assumptions (i)–(iv) are valid for this special function.

First, by straightforward application of the product rule (cf. Nashed [16, p. 165]) to

$$R:\begin{cases}\mathbb{R}^{r\times n}\to\mathbb{R}^{r\times k},\\ C\to CW-S,\end{cases}$$

we see that Q is twice Fréchet-differentiable on $\mathbb{R}^{r\times n}$ and

(i) $\nabla Q[C]\in\mathscr{L}(\mathbb{R}^{r\times n},\mathbb{R}^{k\times k})$ is given by

$$\nabla Q[C](Z)=(ZW)'(CW-S)+(CW-S)'ZW, \tag{33}$$

(ii) $\nabla^2 Q[C]\in\mathscr{L}(\mathbb{R}^{r\times n}\times\mathbb{R}^{r\times n},\mathbb{R}^{k\times k})$ is given by

$$\nabla^2 Q[C](Z_1,Z_2)=(Z_1W)'Z_2W+(Z_2W)'Z_1W. \tag{34}$$

Moreover, for each $C, Z\in\mathbb{R}^{r\times n}$ we have

$$\nabla^2 Q[C](Z,Z)=2W'Z'ZW, \tag{35}$$

which is an n.n.d. matrix. Finally, the inequalities (24) and (25) of assumptions (iii) and (iv) can be seen from

$$\begin{aligned}\|\nabla Q[C](Z)\|^2 &= 2\operatorname{tr}\big([(ZW)'(CW-S)]^2\big)\\ &\quad+2\operatorname{tr}\big((CW-S)'ZW(ZW)'(CW-S)\big)\\ &\leqslant 4\operatorname{tr}\big((CW-S)'ZW(ZW)'(CW-S)\big)\\ &=4\operatorname{tr}\big((ZW)'(CW-S)(CW-S)'ZW\big)\\ &\leqslant 4\lambda_{\max}\big((CW-S)(CW-S)'\big)\operatorname{tr}\big((ZW)'ZW\big)\\ &=4\lambda_{\max}(Q(C))\operatorname{tr}\big((ZW)'ZW\big)\\ &\leqslant 4\lambda_{\max}(Q(C))\lambda_{\max}(WW')\|Z\|^2\end{aligned} \tag{36}$$

and

$$\begin{aligned}\operatorname{tr}\left(\nabla^2 Q[C](Z,Z)\right) &= 2\operatorname{tr}(W'Z'ZW)\\ &= 2\operatorname{tr}(ZWW'Z')\\ &\leqslant 2\lambda_{\max}(WW')\|Z\|^2.\end{aligned} \tag{37}$$

So actually the function Q defined in (32) satisfies assumptions (i)–(iv).

REMARK 1. The inequality (31) now reads as follows:

$$0 \leqslant \nabla^2 g_p[C](Z,Z) \leqslant 2(4p-1)\lambda_{\max}(WW')\|Z\|^2. \tag{38}$$

4. SOLVING THE MINIMAX PROBLEM

We are now in the position to give some general results for solving the optimization problem (12). Note that

$$g:\begin{cases}\mathbb{R}^{r\times n} \to \mathbb{R},\\ C \to \rho\operatorname{tr}(C'C)\end{cases} \tag{39}$$

is uniformly convex and twice Fréchet-differentiable on $\mathbb{R}^{r\times n}$, and

(i) $\nabla g[C] \in \mathscr{L}(\mathbb{R}^{r\times n},\mathbb{R})$ is given by

$$\nabla g[C](Z) = 2\rho\operatorname{tr}(C'Z); \tag{40}$$

(ii) $\nabla^2 g[C] \in \mathscr{L}(\mathbb{R}^{r\times n}\times\mathbb{R}^{r\times n},\mathbb{R})$ is given by

$$\nabla^2 g[C](Z_1,Z_2) = 2\rho\operatorname{tr}(Z_1'Z_2). \tag{41}$$

COROLLARY 1.

(i) *The function* $f_p = g + g_p$ *is uniformly convex and twice Fréchet-differentiable on* $\mathbb{R}^{r\times n}\setminus\mathscr{Q}$.

(ii) *For each* $C \in \mathbb{R}^{r\times n}\setminus\mathscr{Q}$, $Z \in \mathbb{R}^{r\times n}$,

$$2\rho\|Z\|^2 \leqslant \nabla^2 f_p[C](Z,Z) \leqslant \left[2\rho + 2(4p-1)\lambda_{\max}(WW')\right]\|Z\|^2. \tag{42}$$

Proof. (i): Obvious.
(ii): For each $Z \in \mathbb{R}^{r \times n}$ we have

$$\nabla^2 g[C](Z, Z) = 2\rho \|Z\|^2, \tag{43}$$

and thus (42) follows from (43) and (38). ∎

Now let $\tilde{C} \in \mathbb{R}^{r \times n}$ be the solution of

$$\tilde{f}(C) = \rho \operatorname{tr}(C'C) + \operatorname{tr}(Q(C)) \to \min \tag{44}$$

[choose $\tilde{C} = SW'(\rho I + WW')^{-1}$ (cf. Stahlecker and Trenkler [22, Theorem 2])]. Define $\omega = f_1(\tilde{C})$, and consider the level set

$$\mathscr{W}_p(\omega) = \left\{ C \in \mathbb{R}^{r \times n} \middle| f_p(C) \leqslant \omega \right\}. \tag{45}$$

PROPOSITION 6. *The set $\mathscr{W}_p(\omega) \cap \mathscr{Q}$ is empty.*

Proof. Suppose the set $\mathscr{W}_p(\omega) \cap \mathscr{Q}$ is nonempty. Then there exists $\bar{C} \in \mathscr{Q}$ satisfying

$$\tilde{f}(\bar{C}) = f_p(\bar{C}) \leqslant \omega = f_1(\tilde{C}) \leqslant \tilde{f}(\tilde{C}).$$

Since $\tilde{f}$ is uniformly convex on $\mathbb{R}^{r \times n}$, we obtain $\bar{C} = \tilde{C}$. Hence by the first-order condition (cf. Stahlecker and Trenkler [22]) $\rho \tilde{C} = 0$, implying $\tilde{C} = 0$, from which by (5) $Q(\bar{C}) = Q(\tilde{C}) = S \neq 0$ follows, a contradiction. ∎

In view of Proposition 6, the first and second Fréchet derivatives of f_p exist on $\mathscr{W}_p(\omega)$. Therefore first- and second-order methods for solving the auxiliary problems (12) can be used. A general class of procedures will be investigated in the next section.

5. A GLOBALLY CONVERGENT ITERATION PROCEDURE

For given $p \in \mathbb{N}$ consider the iteration procedure

$$C_{t+1} = C_t + \gamma_t S_t, \qquad t = 0, 1, 2, \ldots, \tag{46}$$

where $C_0 \in \mathscr{W}_p(\omega)$ is given and $S_t \in \mathbb{R}^{r\times n}$ is chosen such that

$$\nabla f_p[C_t](S_t) < 0. \tag{47}$$

The stepsize γ_t is chosen according to the stepsize rule of Goldstein (cf. Warth and Werner [25, p. 62]), i.e., γ_t satisfies the inequalities

$$\gamma_t(1-\beta)\nabla f_p[C_t](S_t) \leqslant f_p(C_t+\gamma_t S_t) - f_p(C_t) \tag{48}$$

and

$$f_p(C_t+\gamma_t S_t) - f_p(C_t) \leqslant \gamma_t \beta \nabla f_p[C_t](S_t), \tag{49}$$

where $0<\beta<\frac{1}{2}$. It can be shown that there exists γ_t satisfying (48) and (49) (see Ortega and Rheinboldt [17, p. 257] for the general case). Combining (47) and (49), we see that $f_p(C_{t+1}) \leqslant f_p(C_t)$ for all $t \in \mathbb{N}$. Hence $C_t \in \mathscr{W}_p(\omega)$ for all $t \in \mathbb{N}$.

Note that by (48) and Lemma 4 of the appendix,

$$\begin{aligned} f_p(C_t) + \gamma_t(1-\beta)\nabla f_p[C_t](S_t) &\leqslant f_p(C_t+\gamma_t S_t) \\ &\leqslant f_p(C_t) + \gamma_t \nabla f_p[C_t](S_t) + \frac{M_p}{2}\gamma_t^2 \|S_t\|^2, \end{aligned}$$

where

$$M_p := 2\rho + 2(4p-1)\lambda_{\max}(WW'). \tag{50}$$

For $\gamma_t > 0$ it follows that

$$-\beta \nabla f_p[C_t](S_t) \leqslant \frac{M_p}{2}\gamma_t \|S_t\|^2. \tag{51}$$

Furthermore, by Lemma 2(i) of the appendix

$$f_p(C_t+\gamma_t S_t) - f_p(C_t) \geqslant \gamma_t \nabla f_p[C_t](S_t) + \frac{m}{2}\gamma_t^2 \|S_t\|^2,$$

where $m := 2\rho$, and hence by (49) (since $\gamma_t > 0$)

$$\frac{m}{2}\gamma_t\|S_t\|^2 \leqslant (\beta - 1)\nabla f_p[C_t](S_t).$$

Let now θ_t be defined by

$$\theta_t := \frac{\nabla f_p[C_t](S_t)}{\|F_p(C_t)\|\|S_t\|}, \tag{52}$$

where $F_p : \mathbb{R}^{r\times n} \to \mathbb{R}^{r\times n}$ denotes the gradient of f_p. Then, using (49) and (51),

$$\begin{aligned} f_p(C_t) - f_p(C_t + \gamma_t S_t) &\geqslant -\gamma_t\beta\nabla f_p[C_t](S_t) \\ &\geqslant \frac{2\beta^2}{M_p}\theta_t^2\|F_p(C_t)\|^2. \end{aligned} \tag{53}$$

Furthermore, by Lemma 3 of the appendix we derive from (53)

$$f_p(C_t) - f_p(C_{t+1}) \geqslant c_p\theta_t^2\left[f_p(C_t) - f_p(C_p)\right],$$

where

$$c_p := \frac{4\beta^2 m}{M_p}. \tag{54}$$

Thus

$$f_p(C_{t+1}) - f_p(C_p) \leqslant \left(1 - c_p\theta_t^2\right)\left[f_p(C_t) - f_p(C_p)\right],$$

yielding

$$f_p(C_{t+1}) - f_p(C_p) \leqslant \left(\prod_{i=0}^{t}\left(1 - c_p\theta_i^2\right)\right)\left[f_p(C_0) - f_p(C_p)\right].$$

Hence it follows by Lemma 3 of the appendix that

$$\begin{aligned}\frac{m}{2}\|C_{t+1}-C_p\|^2 &\leqslant f_p(C_{t+1})-f_p(C_p)\\ &\leqslant \left(\prod_{i=0}^{t}\left(1-c_p\theta_i^2\right)\right)\left[f_p(C_0)-f_p(C_p)\right]\\ &\leqslant \left(\prod_{i=0}^{t}\left(1-c_p\theta_i^2\right)\right)\frac{1}{2m}\|F_p(C_0)\|^2.\end{aligned}$$

We may now prove the following theorem.

THEOREM 1. *Consider the iteration procedure* (46). *Assume that* (47) *is valid for all* $t \in \mathbb{N}$ *and the stepsize* γ_t *satisfies the conditions* (48) *and* (49). *Furthermore, the Zoutendijk condition* $\sum_{i=0}^{\infty}\theta_i^2=\infty$ *is assumed to be satisfied. Then*

$$\lim_{t\to\infty} C_t = C_p, \tag{55}$$

and the error estimates

$$\begin{aligned}f_p(C_{t+1})-f_p(C_p) &\leqslant \left(\prod_{i=0}^{t}\left(1-c_p\theta_i^2\right)\right)\left[f_p(C_0)-f_p(C_p)\right]\\ &\leqslant \left(\prod_{i=0}^{t}\left(1-c_p\theta_i^2\right)\right)\frac{1}{2m}\|F_p(C_0)\|^2\end{aligned} \tag{56}$$

and

$$\|C_{t+1}-C_p\| \leqslant \left(\prod_{i=0}^{t}\left(1-c_p\theta_i^2\right)\right)^{1/2}\frac{1}{m}\|F_p(C_0)\| \tag{57}$$

are valid.

Proof. The convergence of the sequence $(C_t)_{t\in\mathbb{N}}$ follows easily, since

$$\prod_{i=0}^{t}\left(1-c_p\theta_i^2\right)\leqslant \exp\left(-c_p\sum_{i=0}^{t}\theta_i^2\right)$$

and $\lim_{t\to\infty}\sum_{i=0}^{t}\theta_i^2=\infty$. ∎

By Theorem 1 the iteration procedure (46) yields a sequence $(C_t)_{t\in\mathbb{N}}$ converging to the unique solution C_p of (12) for each starting point $C_0\in\mathscr{W}_p(\omega)$. We only have to assure that the condition

$$\sum_{i=0}^{\infty}\theta_i^2=\infty \tag{58}$$

is satisfied.

REMARK 2. It can be shown that the class of iteration procedures satisfying (58) contains for instance the steepest-descent method, all restart versions of the conjugate-gradient method, the BFGS method, and the damped Newton method (see for instance Dennis and Schnabel [3] and Ortega and Rheinboldt [17]).

THEOREM 2.

(i) *Let* $\hat{C}_p$ *denote a computed solution of* (12). *Then for each* $C\in\mathbb{R}^{r\times n}$ *the error estimates*

$$\begin{aligned}\|\hat{C}_p-C_*\|&\leqslant\|\hat{C}_p-C_p\|+\|C_p-C_*\|\\&\leqslant\frac{1}{m}\left\|F_p\left(\hat{C}_p\right)\right\|+\left(\frac{2}{m}(k^{1/2p}-1)f(C)\right)^{1/2}\end{aligned}\tag{59}$$

and

$$\begin{aligned}f\left(\hat{C}_p\right)-f(C_*)&\leqslant f_p\left(\hat{C}_p\right)-f_p(C_p)+f_p(C_p)-f(C_*)\\&\leqslant\frac{1}{2m}\left\|F_p\left(\hat{C}_p\right)\right\|^2+(k^{1/2p}-1)f(C)\end{aligned}\tag{60}$$

are valid.

(ii) *For given* $\delta, \varepsilon > 0$ *there exist* $\hat{C}_{p(\delta)}, \hat{C}_{p(\varepsilon)}$ *such that*

$$\|\hat{C}_{p(\delta)} - C_*\| < \delta$$

and

$$f\left(\hat{C}_{p(\varepsilon)}\right) - f(C_*) < \varepsilon.$$

Proof. (i): (59) and (60) follow from (13), (14) and Lemma 3 of the appendix.

(ii): Given $\delta > 0$, first determine $p(\delta) \in \mathbb{N}$ such that for some $C \in \mathbb{R}^{r \times n}$

$$\left(\frac{2}{m}(k^{1/2p(\delta)} - 1) f(C)\right)^{1/2} < \frac{\delta}{2},$$

and then by Theorem 1 determine $C_{t_{p(\delta)}}$ satisfying

$$\frac{1}{m}\left\| F_p\left(C_{t_{p(\delta)}}\right)\right\| < \frac{\delta}{2}.$$

The second inequality follows analogously. ∎

Observe that in the proof of Theorem 2(ii) $p(\delta)$ may be rather large and hence $(C_t)_{t \in \mathbb{N}}$ converges slowly [see (57)]. However, (57) also depends on $\|F_p(C_0)\|$. Hence it is reasonable to choose an increasing sequence $(p_j)_{j \in \mathbb{N}}$, where $p_j \in \mathbb{N}$, and solve

$$f_{p_j}(C) \to \min$$

sequentially to get a good starting point $C_0 \in \mathscr{W}_p(\omega)$ at every p-step.

6. APPLICATION TO LINEAR MINIMAX ESTIMATION

In this section we will use our results to solve a problem of linear minimax estimation.

Consider a general linear model

$$y \sim (X\beta, V), \tag{61}$$

where y is the n-dimensional random vector of observations with expectation $X\beta$ and covariance matrix V. The $n \times k$ regressor matrix X and the $n \times n$ p.d. covariance matrix V are assumed to be known. Suppose that the unknown parameter vector β belongs to an ellipsoid

$$\mathscr{E} = \left\{\beta \in \mathbb{R}^k \middle| (\beta - \beta_0)'T(\beta - \beta_0) \leqslant 1\right\} \tag{62}$$

and, in addition, satisfies affine linear restrictions of the form

$$\mathscr{L} = \left\{\beta \in \mathbb{R}^k \middle| R\beta = r\right\}, \tag{63}$$

i.e. $\beta \in \mathscr{B}$, where

$$\mathscr{B} = \mathscr{E} \cap \mathscr{L}. \tag{64}$$

Here T is a known $k \times k$ p.d. matrix, β_0 a known $k \times 1$ vector, R a known $m \times k$ matrix, and r a known $k \times 1$ vector. From now on we assume that $\mathscr{B}$ is nonempty.

Let $B\beta$ be a linear function of β to be estimated, where B is a given $q \times k$ matrix, and let

$$\mathscr{D} = \left\{d \middle| d = B[c + C(y - Xc)],\ C \in \mathbb{R}^{k \times n},\ c \in \mathbb{R}^k\right\} \tag{65}$$

be a class of affine linear statistics for $B\beta$. Finally, consider the weighted mean squared error of a d from this class, given by

$$\begin{aligned} R_A(d, B\beta) &= E[(d - B\beta)'A(d - B\beta)] \\ &= \operatorname{tr}(VC'AC) + (\beta - c)'(CX - I)'A(CX - I)(\beta - c). \end{aligned} \tag{66}$$

An estimator $d_* \in \mathscr{D}$ is called a minimax inhomogeneous linear estimator (MILE) for $B\beta$ iff d_* minimizes

$$\max_{\beta \in \mathscr{B}} R_A(d, B\beta) \tag{67}$$

over $d \in \mathscr{D}$, i.e. iff the function

$$\operatorname{tr}(VC'AC) + \max_{\beta \in \mathscr{B}} (\beta - c)'(CX - I)'A(CX - I)(\beta - c) \tag{68}$$

is minimized over $C \in \mathbb{R}^{k \times n}$ and $c \in \mathbb{R}^k$.

The problem of obtaining a MILE under the prior information $\beta \in \mathscr{B}$ has recently been investigated by Stahlecker and Trenkler [22] and Drygas [5]. Previous results on MILE for the case of ellipsoidal constraints can be found in Kuks and Olman [12], Läuter [13], and Hoffmann [10]. Starting from the case $\beta \in \Omega$, where $\Omega \subset \mathbb{R}^k$ is a symmetric compact set, and developing connections to the theory of Bayes-optimal design, recently Pilz [18], Gaffke and Heiligers [6], and Gaffke and Mahtar [8] investigated the problem of determining linear minimax estimators, where specific attention is also paid to the case of ellipsoidal restrictions.

Following Stahlecker and Trenkler [22], we define

$$S = RT^{-1/2},$$

$$s = r - R\beta_0,$$

$$t_* = S^+ s,$$

$$\alpha = 1 - t_*' t_*,$$

$$P = I - S^+ S,$$

and

$$G = B'AB.$$

It can be shown by a symmetry argument that a minimax initial guess for c is given by

$$c_* = \beta_0 + T^{-1/2} t_*. \tag{69}$$

From Theorem 1 of Stahlecker and Trenkler [22] we now conclude that a MILE $d_* \in \mathscr{D}$ may be obtained by minimizing

$$\mathrm{tr}(VC'GC) + \lambda_{\max}(Q(C)) \tag{70}$$

over $C \in \mathbb{R}^{k \times n}$, where

$$Q(C) = \alpha P T^{-1/2}(CX - I)' G (CX - I) T^{-1/2} P. \tag{71}$$

Obviously, this problem is one of the class (1) which we studied in the sections before. Setting

$$d_p = B\left[c_* + C_p(y - Xc_*)\right], \tag{72}$$

where $C_p \in \mathbb{R}^{k\times n}$ solves the auxiliary problem

$$f_p(C) \to \min \tag{73}$$

for $p \in \mathbb{N}$, we get an approximate MILE. Since we know that $\lim_{p\to\infty} C_p = C_*$, where $C_* \in \mathbb{R}^{k\times n}$ is a solution of (70), it follows that $d_p \to d_*$ almost surely as $p \to \infty$.

APPENDIX

Lemma 1. *Let $Q : \mathbb{R}^{r\times n} \to \mathbb{R}^{k\times k}_{\mathscr{S}}$ be an arbitrary twice Fréchet-differentiable matrix function. For each $C, D \in \mathbb{R}^{r\times n}$ and $\mu \in [0,1]$ we have*

$$\mu Q(C) + (1-\mu)Q(D) - Q(\mu C + (1-\mu)D) \in \mathbb{R}^{k\times k}_{\geqslant}$$

(i.e., Q is convex with respect to the Löwner ordering on $\mathbb{R}^{k\times k}_{\geqslant}$) if and only if for all $C, Z \in \mathbb{R}^{r\times n}$ one has $\nabla^2 Q[C](Z,Z) \in \mathbb{R}^{k\times k}_{\geqslant}$.

Proof. For any $y \in \mathbb{R}^k$ define $f_y : \mathbb{R}^{r\times n} \to \mathbb{R}$ by $f_y(C) := \operatorname{tr}(yy'Q(C))$. Observe that Q is convex iff f_y is convex for any $y \in \mathbb{R}^k$. The statement now follows from

$$\nabla^2 f_y[C](Z,Z) = \operatorname{tr}\left(yy'\nabla^2 Q[C](Z,Z)\right). \qquad \blacksquare$$

From now on let $\mathscr{V}$ be a Hilbert space and $\mathscr{D} \subseteq \mathscr{V}$ convex.

Definition 1. A real-valued functional g defined on $\mathscr{D}$ is said to be uniformly convex if there exists $\delta > 0$ such that for all $x, y \in \mathscr{D}$ and $\lambda \in [0,1]$

$$\lambda g(x) + (1-\lambda)g(y) - g(\lambda x + (1-\lambda)y) \geqslant \frac{\delta}{2}\lambda(1-\lambda)\|x-y\|^2. \tag{74}$$

DEFINITION 2. An operator $G: \mathscr{D} \to \mathscr{V}$ is said to be

(i) uniformly monotone if there exists $\delta > 0$ such that for all $x, y \in \mathscr{D}$

$$\langle G(x) - G(y), x - y \rangle \geqslant \delta \|x - y\|^2; \tag{75}$$

(ii) monotone if for all $x, y \in \mathscr{D}$

$$\langle G(x) - G(y), x - y \rangle \geqslant 0. \tag{76}$$

In the sequel let $\mathscr{D} \subseteq \mathscr{V}$ be open and convex, $g: \mathscr{D} \to \mathbb{R}$ Fréchet-differentiable, and $G: \mathscr{D} \to \mathscr{V}$ the gradient of g.

In the next lemma we will present some known results on (uniformly) convex functionals (see for instance Göpfert [9, p. 171 and p. 173] and Vainberg [24, pp. 51, 121]).

LEMMA 2. *The functional g is*

(i) *uniformly convex if and only if there exists* $\delta > 0$ *such that for all* $x, y \in \mathscr{D}$

$$\frac{\delta}{2}\|y - x\|^2 \leqslant g(y) - g(x) - \langle G(x), y - x \rangle; \tag{77}$$

(ii) *(uniformly) convex if and only if G is (uniformly) monotone.*

If g is uniformly convex on $\mathscr{D}$, then by Lemma 2(ii) the solution of the optimization problem

$$g(x) \overset{x \in \mathscr{D}}{\to} \min \tag{78}$$

is unique.

LEMMA 3. *Let* $x_* \in \mathscr{D}$ *be the unique solution of* (78). *Then for each* $x \in \mathscr{D}$ *the inequalities*

$$\frac{\delta}{2}\|x - x_*\|^2 \leqslant g(x) - g(x_*) \leqslant \frac{1}{2\delta}\|G(x)\|^2 \tag{79}$$

are valid.

Proof. The first inequality follows by direct application of Lemma 2(i) [note that by the definition of x_*, $\langle G(x_*), x - x_*\rangle = \nabla g[x_*](x - x_*) = 0$].

Furthermore, by Lemma 2(i), for each $h \in \mathscr{V}$ with $x + h \in \mathscr{D}$

$$g(x) + \langle G(x), h\rangle + \frac{\delta}{2}\|h\|^2 \leqslant g(x+h) \tag{80}$$

and hence

$$\min_{h\in\mathscr{V}}\left\{ g(x) + \langle G(x), h\rangle + \frac{\delta}{2}\|h\|^2 \right\} \leqslant \min_{h\in\mathscr{V}} g(x+h) = g(x_*). \tag{81}$$

The left-hand side attains its minimum (cf. Göpfert [9, p. 177] and Allwright [1, p. 370]) for $h_* = -(1/\delta)G(x)$. Thus,

$$g(x) - g(x_*) \leqslant \frac{1}{\delta}\|G(x)\|^2 - \frac{1}{2\delta}\|G(x)\|^2 = \frac{1}{2\delta}\|G(x)\|^2 \tag{82}$$

which proves the second inequality. ■

LEMMA 4. *If g is twice Fréchet-differentiable on $\mathscr{D}$ and there exists $M > 0$ such that for all $x \in \mathscr{D}$ and $h \in \mathscr{V}$*

$$\nabla^2 g[x](h, h) \leqslant M\|h\|^2, \tag{83}$$

then for all $x, y \in \mathscr{D}$

$$\langle G(x) - G(y), x - y\rangle \leqslant M\|x - y\|^2 \tag{84}$$

and

$$g(x) - g(y) - \langle G(y), x - y\rangle \leqslant \frac{M}{2}\|x - y\|^2. \tag{85}$$

Proof. Since (83) is satisfied, (84) follows trivially by the mean-value theorem for operators (cf. Wouk [26, p. 265]). The proof of (85) follows analogously to Göpfert [9, p. 171 and p. 173]. ■

The authors wish to thank the unknown referees for their critical remarks and helpful proposals.

REFERENCES

1 J. C. Allwright, Global convergence rates for descent algorithms: Improved results for strictly convex problems, *J. Optim. Theory Appl.* 26:367–372 (1978).
2 J. W. Daniel, *The Approximate Minimization of Functionals*, Prentice-Hall, Englewood Cliffs, N.J., 1971.
3 J. E. Dennis and R. B. Schnabel, *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*, Prentice-Hall, Englewood Cliffs, N.J., 1983.
4 H. Drygas, Minimax prediction in linear models, in *Linear Statistical Inference* (W. Klonecki and T. Caliński, Eds.), Springer-Verlag, New York, 1982, pp. 48–60.
5 H. Drygas, A Note on Minimax Estimation in Regression Models with Affine Restrictions, Technical Report 2/88, Mathematische Schriften der Gesamthochschule Kassel, 1988.
6 N. Gaffke and B. Heiligers, Bayes, Admissible, and Linear Minimax Estimators in Linear Models with Restricted Parameter Space, Technical Report 132, Univ. of Augsburg, Inst. of Mathematics, 1987.
7 N. Gaffke and O. Krafft, Matrix inequalities in the Löwner ordering, in *Modern Applied Mathematics* (B. Korte, Ed.), North-Holland, New York, 1982, pp. 595–622.
8 N. Gaffke and R. Mahtar, Linear Minimax Estimation under Ellipsoidal Parameter Space and Related Bayes *L*-Optimal Designs, Technical Report 42, Schwerpunktprogramm der Deutschen Forschungsgemeinschaft, Anwendungsbezogene Optimierung und Steuerung, 1988.
9 A. Göpfert, *Mathematische Optimierung in allgemeinen Vektorräumen*, Teubner, Leipzig, 1973.
10 K. Hoffmann, Characterization of minimax linear estimators in linear regression, *Math. Operationsforschung und Statist. Ser. Statist.* 10:19–26 (1979).
11 R. A. Horn and C. A. Johnson, *Matrix Analysis*, Cambridge U.P., Cambridge, 1985.
12 J. Kuks and W. Olman, Minimax linear estimation of regression coefficients II (in Russian), *Izv. Akad. Nauk Eston. SSR* 21:66–72 (1972).
13 H. Läuter, A minimax linear estimator for linear parameters under restrictions in form of inequalities, *Math. Operationsforschung und Statist. Ser. Statist.* 6:689–695 (1975).
14 J. R. Magnus and H. Neudecker, *Matrix Differential Calculus with Applications in Statistics and Econometrics*, Wiley, Chichester, 1988.
15 A. W. Marshall and I. Olkin, *Inequalities: Theory of Majorization and Its Applications*, Academic, New York, 1979.
16 M. Z. Nashed, Differentiability and related properties of nonlinear operators: Some aspects of the role of differentials in nonlinear functional analysis, in *Nonlinear Functional Analysis and Applications* (L. B. Rall, Ed.), Academic, New York, 1971, pp. 103–309.
17 J. M. Ortega and W. C. Rheinboldt, *Iterative Solution of Nonlinear Equations in Several Variables*, Academic, New York, 1970.

18 J. Pilz, Minimax linear regression estimation with symmetric parameter restriction, *J. Statist. Plann. Inference* 13:297–318 (1986).
19 J. Pilz, Beiträge zur Theorie der Bayes- und minimax-linearen Schätzungen in linearen Regressionsmodellen (in German), D. Sc. Thesis, Bergakademie Freiberg, 1987.
20 P. Stahlecker, *A priori Information und Minimax-Schätzung im Linearen Regressionsmodell* (in German), Math. Systems in Econom. 108, Athenäum, Frankfurt, 1987.
21 P. Stahlecker and J. Lauterbach, Approximate linear minimax estimation in regression analysis with ellipsoidal constraints, *Comm. Statist. Theory Methods* 18:2755–2784 (1989).
22 P. Stahlecker and G. Trenkler, Full and partial minimax estimation in regression analysis with additional linear constraints, *Linear Algebra Appl.* 111:279–292 (1988).
23 H. Toutenburg, *Prior Information in Linear Models*, Wiley, New York, 1982.
24 M. M. Vainberg, *Variational Method and Method of Monotone Operators in the Theory of Nonlinear Equations*, Wiley, New York, 1973.
25 W. Warth and J. Werner, Effiziente Schrittweitenfunktionen bei unrestringierten Optimierungsproblemen, *Computing* 19:59–72 (1977).
26 A. Wouk, *A Course of Applied Functional Analysis*, Wiley, New York, 1979.

Received 4 October 1988; final manuscript accepted 29 March 1989